

RESEARCH ARTICLE

# A Bayesian Developmental Approach to Robotic Goal-Based Imitation Learning

Michael Jae-Yoon Chung<sup>1</sup>, Abram L. Friesen<sup>1</sup>, Dieter Fox<sup>1</sup>, Andrew N. Meltzoff<sup>2</sup>, Rajesh P. N. Rao<sup>1\*</sup>

**1** Department of Computer Science & Engineering, University of Washington, Seattle, WA, United States of America, **2** Institute for Learning & Brain Sciences, University of Washington, Seattle, WA, United States of America

\* [rao@cs.washington.edu](mailto:rao@cs.washington.edu)



**OPEN ACCESS**

**Citation:** Chung MJ-Y, Friesen AL, Fox D, Meltzoff AN, Rao RPN (2015) A Bayesian Developmental Approach to Robotic Goal-Based Imitation Learning. PLoS ONE 10(11): e0141965. doi:10.1371/journal.pone.0141965

**Editor:** Rachel L Kendal, Centre for Coevolution of Biology & Culture, University of Durham, UNITED KINGDOM

**Received:** January 22, 2015

**Accepted:** October 15, 2015

**Published:** November 4, 2015

**Copyright:** © 2015 Chung et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data and code for this paper are now available on github as follows:

Robot data: [https://github.com/mjyc/chung2015plosone\\_robot\\_data](https://github.com/mjyc/chung2015plosone_robot_data). Matlab code: <https://github.com/afriesen/gpgaze>.

**Funding:** This work was supported by Office of Naval Research (ONR) Science of Autonomy, grant no. N000141310817, <http://www.onr.navy.mil/>; Office of Naval Research (ONR) Cognitive Science program, grant no. N000140910097, <http://www.onr.navy.mil/>; National Science Foundation (NSF), NSF grant no. SMA-1540619 and NSF grant no. IIS-

## Abstract

A fundamental challenge in robotics today is building robots that can learn new skills by observing humans and imitating human actions. We propose a new Bayesian approach to robotic learning by imitation inspired by the developmental hypothesis that children use self-experience to bootstrap the process of intention recognition and goal-based imitation. Our approach allows an autonomous agent to: (i) learn probabilistic models of actions through self-discovery and experience, (ii) utilize these learned models for inferring the goals of human actions, and (iii) perform goal-based imitation for robotic learning and human-robot collaboration. Such an approach allows a robot to leverage its increasing repertoire of learned behaviors to interpret increasingly complex human actions and use the inferred goals for imitation, even when the robot has very different actuators from humans. We demonstrate our approach using two different scenarios: (i) a simulated robot that learns human-like gaze following behavior, and (ii) a robot that learns to imitate human actions in a tabletop organization task. In both cases, the agent learns a probabilistic model of its own actions, and uses this model for goal inference and goal-based imitation. We also show that the robotic agent can use its probabilistic model to seek human assistance when it recognizes that its inferred actions are too uncertain, risky, or impossible to perform, thereby opening the door to human-robot collaboration.

## Introduction

Considerable progress has been made in robotics in recent years, particularly in the area of human-robot interaction. Techniques have been proposed to impart new skills to robots via “programming by demonstration” [1] and “imitation learning” [2]. An important remaining challenge is endowing a robot with the ability to infer the intentions of humans and to learn new skills not by naively following demonstrated action trajectories (“trajectory-based” or “action” imitation) but through *goal-based imitation* [3, 4].

Developmental scientists have shown that infants can infer the goal of an adult’s actions. In one set of experiments with 18-month old infants [5], an adult actor demonstrated an act in

1318733, <http://www.nsf.gov/>; Intel Science and Technology Center (ISTC), Seattle; and (NSERC) fellowship to A. Friesen, [http://www.nserc-crsng.gc.ca/index\\_eng.asp](http://www.nserc-crsng.gc.ca/index_eng.asp). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The Intel Science and Technology Center, Seattle, is included in the list of acknowledgments due to their assistance with the project. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

which the goal-state was not achieved; infants were able to read through the literal body movements to infer the underlying goal and execute the intended act. In other experiments, children employed different means to achieve the same goal as adults. When given a barbell-shaped object, adults used their hands to pull apart the object, but children who could not grasp the end of an oversized barbell used alternative means (e.g., holding one end in both hands and clasping the other end between their knees) to pull it apart. The children thus not only inferred the adult's goal but could also use novel alternative means to achieve the goal.

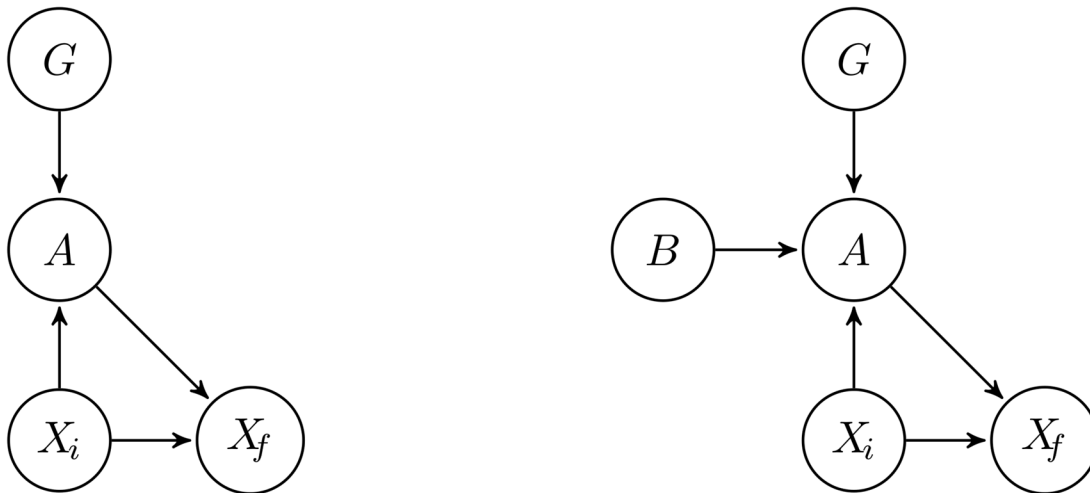
Clues to how children acquire the ability to infer goals have come from studies on gaze following, the ability of humans to follow the line of regard of another human. A naive interpretation of gaze following is that it is simply the imitation of head movement. However, Meltzoff and Brooks [6] controlled for head movements and showed that infants' tendency to follow gaze varied as a function of whether or not the actor's view was blocked by an intervening opaque occluder (a blindfold). In particular, 18-month-old children did not follow the gaze of an adult who made a head movement toward an object while wearing a blindfold. Younger children (12-month-olds) did mistakenly follow the gaze of a blindfolded adult. However, after these younger children were given the self-experience of a blindfold blocking their own view, they no longer followed the gaze of the blindfolded adult.

These results highlight the importance of self-experience in the development of goal inference capabilities and goal-based imitation. They are consistent with Meltzoff's "Like-Me" hypothesis [7] which states that children utilize internal models learned through self-experience to interpret the acts and intentions of others, and with increasing experience, acquire increasingly sophisticated intent recognition abilities. Such an approach is different from but incorporates elements of two previous well-known theories of intent inference: (i) "simulation theory" [8], which proposes that the same mental resources we use for thinking, decision-making, and emotional responses are redeployed in imagination to provide an understanding of others, and (ii) "theory theory" [9, 10], which advocates that children develop theories about the world and about others, make predictive inferences about behaviors and inner states of others using this network of beliefs ("theories"), and revise theories according to new observations. The approach taken here differs by providing a rigorous mathematical framework based on Bayesian inference of goals and actions, and the bootstrapping of Bayesian models through learning from self-experience.

## Methods

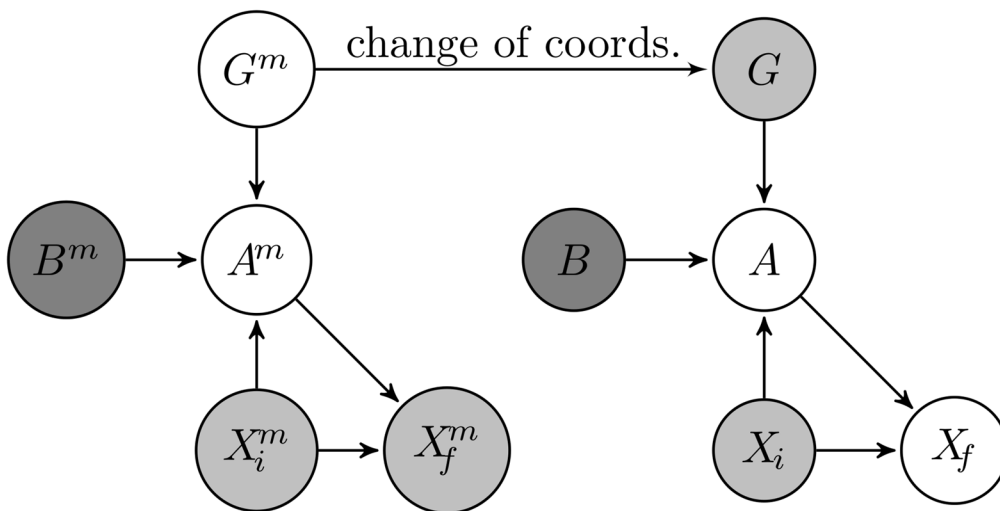
Our hypothesis is that humans use a goal-directed mechanism for planning motor acts. A goal, provided by either an internal desire or external stimulus, together with the current state, determines what the action should be—this is called a "policy." An action executed in a given state in turn determines the next state (probabilistically). We represent these dependencies using the graphical model shown in Fig 1a where  $G$  is the goal,  $A$  is the action,  $X_i$  is the current (or initial) state, and  $X_f$  is the next (or final) state. Throughout this paper, we use upper case letters to denote random variables and lower case letters to denote specific values of the variables.

Consider our two example scenarios. In the context of gaze following, the goal is a desired fixation location, the action is a vector of motor commands to move the head, and the state represents head position and orientation. We discuss the model assuming movement of the head but the model can be generalized to embody movement of the eyes or movement of both the head and eyes to achieve a spatial target. In this case, the random variables (states, actions, goals) are continuous-valued and we define the relationships between them using (learned) *probabilistic functions*. In the case of our example of actions on objects, we assume the states represent discrete locations on a table, actions comprise of high-level commands such as pick



(a) Basic model

(b) Model with blindfold experience



(c) Gaze-following model

**Fig 1. Graphical models for gaze following.** (a) The graphical model shows the probabilistic dependencies between different random variables in the model:  $G$  = goal,  $A$  = action,  $X_i$  = current state, and  $X_f$  = final state. The model captures how actions depend on goals and states, and how the state changes as a consequence of executing an action; (b) incorporates the influence of blindfold self-experience on the model using the random variable  $B$ ; (c) shows the combined graphical models, one for the agent and a copy for the mentor (superscript  $m$ ), for following the gaze of a mentor. Shaded variables denote observed variables. The darker shading indicates that  $B$  is an observed discrete variable, while the rest of the nodes are continuous.

doi:10.1371/journal.pone.0141965.g001

and push, and the goal is a desired (discrete) location on the table. We use this example to illustrate the discrete formulation of the model in which the relationships between the discrete-valued random variables are captured by (learned) probability tables.

## Computational Model

**Case I: Continuous-Valued Random Variables.** We begin by showing how an agent can learn the mapping between continuous-valued goals, states, and actions which can later be used for planning and intent inference. The agent first learns a transition model  $f$ , (e.g., through self-exploration using its own body movements, which [11] dubbed “body babbling”). This translates an initial state,  $x_i$ , and an action,  $a$ , to a final state,  $x_f$ :

$$x_f = f(x_i, a) + v, \quad v \sim \mathcal{N}(0, \Sigma_v),$$

where  $\mathcal{N}(\mu, \Sigma)$  signifies the normal distribution with mean  $\mu$  and covariance matrix  $\Sigma$ .

To make the exposition concrete, we focus here on the specific case of modeling gaze following, although the framework is general enough to apply to other types of behaviors. In gaze following, the initial and final state correspond to head poses while the action corresponds to motor commands to move the head. The learned transition model can in turn be used to learn a “policy” function,  $\pi$ , which maps an initial state and a goal location  $g$  to an appropriate action:

$$a = \pi(g, x_i) + v, \quad v \sim \mathcal{N}(0, \Sigma_v)$$

This equation essentially determines the rotation required to turn the head from its current position to face the goal.

We use two separate Gaussian processes (GPs) [12],  $GP_f$  and  $GP_m$ , to learn the two nonlinear functions  $f$  and  $\pi$  in the equations above. GPs are commonly used in machine learning to infer a function  $h$  from noisy observations  $y_i$ :

$$y_i = h(x_i) + v, \quad v \sim \mathcal{N}(0, \sigma_v^2)$$

GPs are both flexible and robust. First, they are nonparametric so they do not limit the set of functions that  $h$  can be chosen from. Second, they estimate a probability distribution over functions instead of choosing a single most-likely function. This allows all plausible functions to be incorporated into the estimation process and reduces model bias. For these reasons, GPs are effective with small numbers of training samples, increasing their biological plausibility. We now show how these GPs can be used for planning, goal inference, and gaze-following.

**Goal-directed Planning.** GPs are trained via supervised learning: given a training dataset with noisy labels, the GP learns the functional mapping between the input data and the labels (output). The training data could be obtained, for instance, through a reinforcement-based paradigm that combines exploration of the goal-action-state space for training the transition GP with selection of data from successful trials for training the policy GP (see [3]).

After training, it is simple for the agent to fixate on a goal location. We assume that the agent knows its current state  $x_i$  and the goal  $g$ . Given these, the agent uses its learned distribution over functions,  $GP_m$ , to compute a probability distribution over actions,  $P(a) = \mathcal{N}(\mu_a, \Sigma_a)$ . This distribution can then be passed through  $GP_f$  to estimate the probability of the resulting state  $P(x_f) \approx \mathcal{N}(\mu_{x_f}, \Sigma_{x_f})$ . Thus, our model provides both a prediction of the final position and an estimate of the uncertainty in the prediction. We define the combined inference process  $P(X_f | X_i = x_i, G = g)$ , as *Forward Inference*.

**Goal Inference.** One can also infer the goal of an observed action (e.g., head movement), given starting and ending states (e.g., head poses),  $x_i$  and  $x_f$ , respectively. To accomplish this, the agent must be able to recover the inputs to each GP, given the outputs. Fortunately, results

from [13] allow us to estimate a distribution over the inputs given the outputs. We follow the technique in [13] to infer a distribution over actions given  $x_i$  and  $x_f$  and then use this to estimate a distribution over goals. We define the inference process  $P(G|X_i = x_i, X_f = x_f)$  as *Reverse Inference*.

**Gaze Following.** Gaze following of a mentor using the above model is accomplished by invoking Meltzoff's "Like-Me" hypothesis. The agent learns a model of its own head movements and assumes that a mentor uses this same (or similar) model. As suggested by the graphical model in Fig 1c, the agent observes the starting and ending states (head poses) of a mentor and then infers the goal location indicated by the mentor,  $g^m$ , by using a copy of its own learned model and inferring what it would be looking at if it were in the mentor's position. After inferring the mentor's goal, the agent transforms that goal into its own coordinate frame and then infers how to fixate that goal. For this paper, we assume the agent has acquired the ability to transform between coordinate frames through prior experience. We acknowledge that the problem of transformation between the agent and mentor coordinate frames can be challenging in more complex tasks, and we refer the reader to relevant work in this area for further information [14–16].

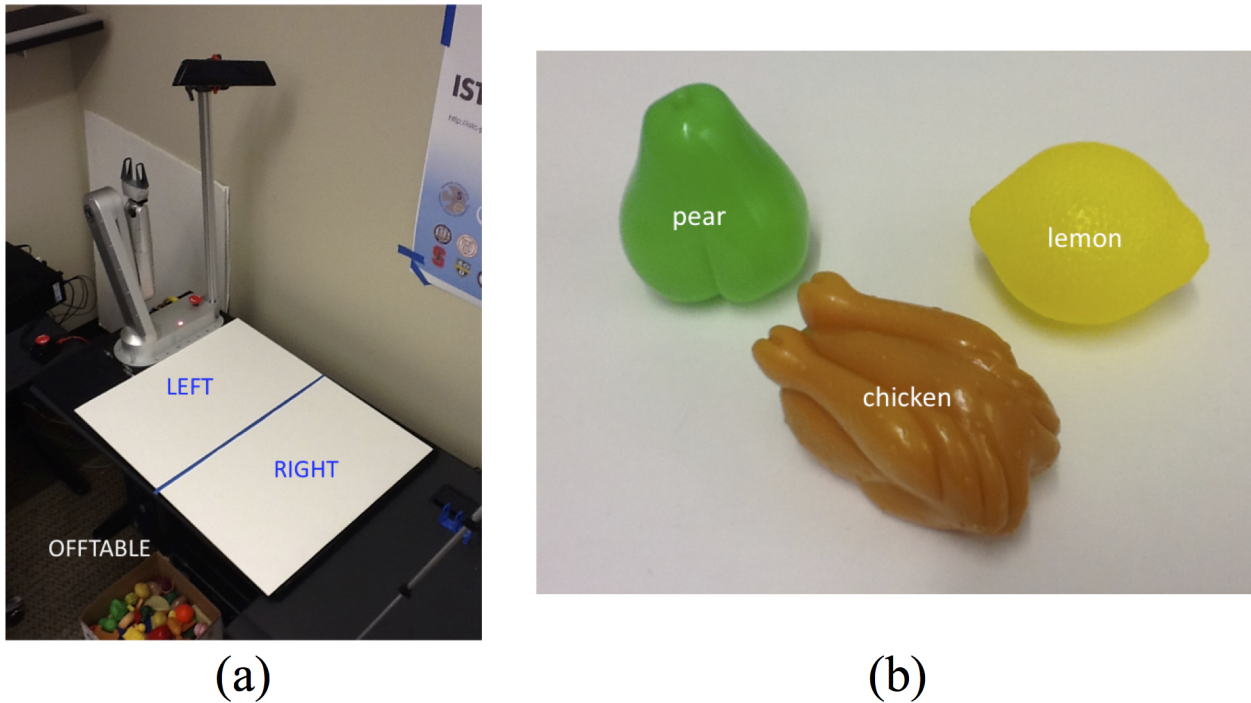
**Modeling Blindfold Experiments in Human Infants.** We first tested the continuous-valued version of the model on a gaze-following task previously used in experiments in human infants [6]. One set of experiments showed that 14- and 18-month olds do not follow the gaze of an adult who is wearing a blindfold, although they follow gaze if the adult wears the same band as a headband. This suggests that these children did not follow gaze because they are able to take into account the consequences of wearing a blindfold (i.e., occlusion) and unlike the 12-month olds, make the inference that the adult is not looking at an object. This observation is closely related to Meltzoff's "Like me" hypothesis [7]. In particular, self-experience with own eye closure and occluders may influence gaze-following behavior. To test this hypothesis, Meltzoff and Brooks provided one group of 12-month olds with self-experience with an opaque blindfold while two other groups either had no self-experience or had self-experience with a windowed blindfold. On seeing an adult with a blindfold turn towards an object, most of the children with self-experience with blindfolds did not turn to the object while the other two groups did [6]. This highlights self-experience as a learning mechanism that can be used to interpret the behaviors of other agents.

To model these results, we incorporate a binary random variable  $B \in \{0,1\}$  in the graphical model, as shown Fig 1b, which denotes whether or not a blindfold is being worn, and allows the agent to learn the effects of being blindfolded. Our model learns a new Gaussian process ( $GP_\pi^B$  in place of  $GP_\pi$ ) which is used when the agent is blindfolded. When the opaque blindfold is in place, regardless of the current value of the goal state, no action leads to that goal location being fixated. Goals in this case are not causally linked to states (head poses) or actions. The agent can learn this and then apply this knowledge to a mentor agent to infer that the mentor does not have a goal when blindfolded (i.e., the inferred distribution over goals approximates a uniform random distribution). However, if this alternate Gaussian process is not learned, the agent does not know the consequences of wearing a blindfold and follows the mentor's head movement even if the mentor is blindfolded. Fig 1c shows the combined graphical model for following the gaze of the mentor, based on combining a model for the agent and a copy for the mentor.

## Case II: Discrete-Valued Random Variables

When the random variables for states, actions, and goals are discrete-valued, the model can be expressed in terms of conditional probability tables for the transition probabilities  $P(X_f|X_i, A)$  and the "policy"  $P(A|G, X)$ .





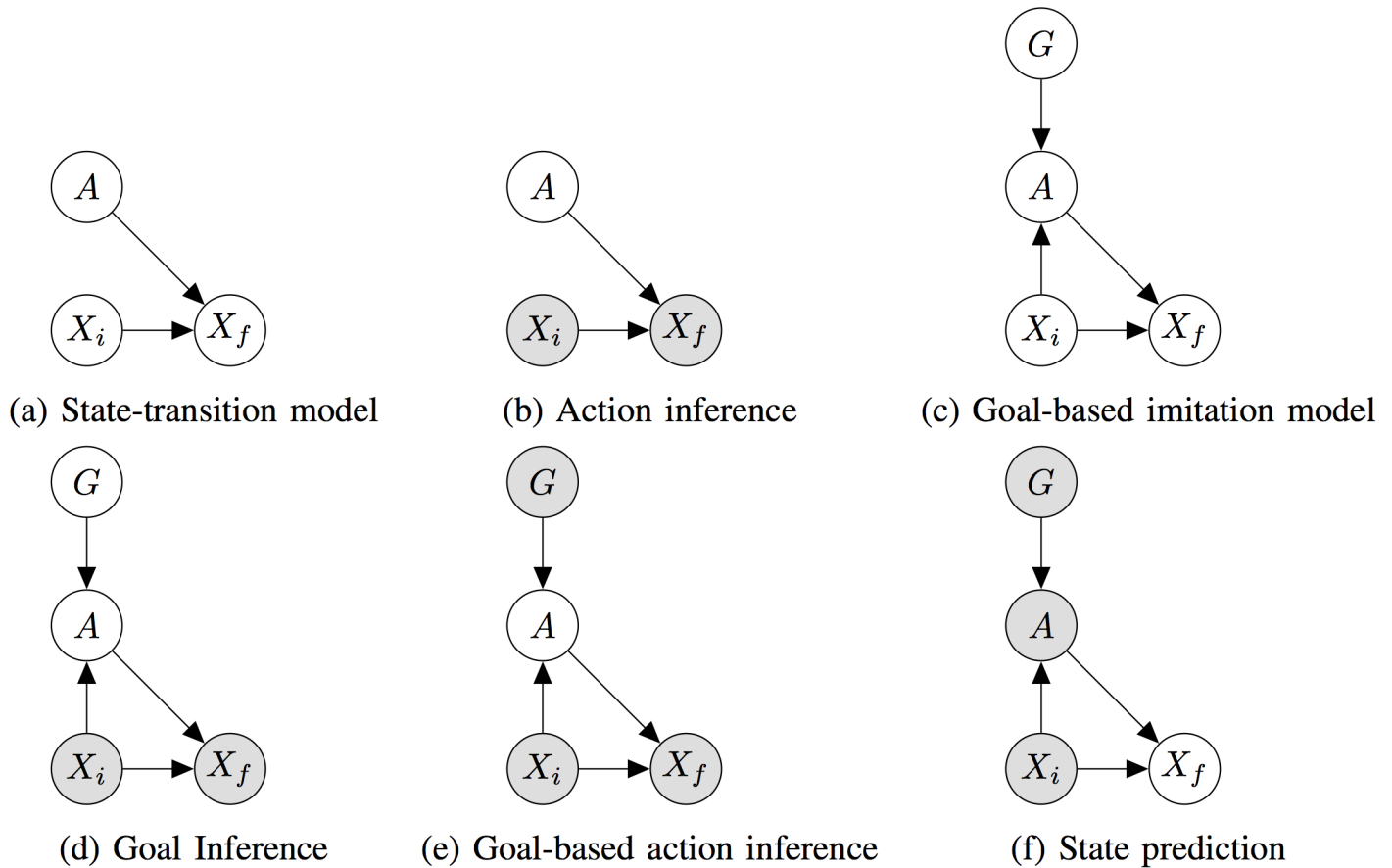
**Fig 2. Robotic tabletop organization task setup.** (a) The robot is located on the left side of the work area and the Kinect looks down from the left side from the robot perspective. The three predefined areas that distinguish object states are notated. (b) Toy tabletop objects.

doi:10.1371/journal.pone.0141965.g002

For concreteness, we describe this case in the context of a simple tabletop task involving a set of small objects on a tabletop which can be moved around by a human or a robotic arm as shown in Fig 2. The discretized position of an object defines its *state* and a *goal* corresponds to the object reaching a particular state. The robotic arm can manipulate the state of any object using a set of *actions*. We define  $\Omega^X$  to be the set of discrete states in the environment,  $\Omega^A$  to be the set of all possible actions available to the robot (these can be different from the possible actions of the human demonstrator) and  $\Omega^G$  to be the set of possible goal states. We assume all three sets are finite.

Each goal  $g \in \Omega^G$  represents an abstract task which can be achieved using one or more actions in  $\Omega^A$ . For example, a goal can be moving an object to location A regardless of its starting location. The object could be picked and placed at A, pushed to A, or transported using some other action, but the goal remains the same. The dynamics of the action and its effect are modeled as a Markov state-transition model (Fig 3a); when the robot in a state  $x_i$  executes an action  $a_i$ , it enters a state  $x_f$  with the transition probability  $P(X_f = x_f | X_i = x_i, A = a_i)$ .

**Learning Through Self-Experience.** We assume that the transition probability distribution is initially unknown to the robot and must be learned through exploration, similar to the manner in which infants explore the consequences of their actions through exploration and “body babbling” [11, 17]. The robot collects training data tuples  $(x, a, x')$  for the graphical model in Fig 3a by executing a random action  $a$  from a random initial state  $x$  and observing a resulting final state  $x'$  multiple times. Given this training data, maximum likelihood parameter estimation is used to learn parameters for the transition probability distribution  $P(X_f | X_i, A)$ .



**Fig 3. Graphical models for robotic goal-based imitation.** (a) through (f) illustrate the use of graphical models for learning state-transitions, action inference, goal inference, goal-based imitation, and state prediction. Shaded nodes denote observed variables.

doi:10.1371/journal.pone.0141965.g003

**Goals and Goal Inference.** After the transition parameters are learned, a goal-based graphical model,  $\mathcal{G}$ , is created by augmenting the initial model in Fig 3a with a new node  $G$  as shown in Fig 3c. The robot then engages in goal-based exploration in which a goal state  $g$  is chosen at random and for any given initial state  $X_i = x_i$ , the robot performs Bayesian inference to infer  $P(A|X_i = x_i, X_f = g)$ . It samples an action  $A^*$  from this distribution and executes this action. If the goal  $g$  is reached, it increases the “policy” probability  $P(A = A^*|G = g, X_i = x_i)$  by a small amount, and decreases  $P(A = A'|G = g, X_i = x_i)$  for all other actions  $A'$ . An alternate approach, which we follow here, is to directly set  $P(A|G = g, X_i = x_i) = P(A|X_i = x_i, X_f = g)$ . In this case, the policy is based purely on the plan inferred from a learned transition model without verification that a goal can be reached, so the policy’s accuracy will depend on the accuracy of the transition model. For the simple tabletop experiment we illustrate here, we adopt the latter approach given that accurate transition models can be learned for the small state and action space.

For goal inference, the robot observes object states  $x_i^h$  and  $x_f^h$  (e.g., object locations) from a human demonstration. By invoking the “Like me” hypothesis, the robot uses its goal-based

graphical model to compute the posterior distribution over goals  $G$  given  $x_i^h$ , and  $x_f^h$ , as depicted in Fig 3d (note that the variable  $A$  is marginalized out during goal inference).

**Goal-Based Imitation and Action Selection.** Goal-based imitation is implemented as a two-stage process: (i) the robot infers the likely goal of the human demonstration using the goal-based graphical model described above, and (ii) either executes the action most likely to achieve this goal, or seeks human assistance if the goal is found to be unlikely to be achieved. Specifically, the robot senses the current state  $x_i^r$  using its sensors and infers the human's goal  $g_{MAP}$  by taking the mode of the posterior distribution of  $G$  from the goal-inference step. It then computes the posterior over actions  $A$  as shown in Fig 3e and selects the maximum a posteriori action  $a_{MAP}$ . Since  $a_{MAP}$  is not guaranteed to succeed, the robot predicts the probability of reaching the most probable final state  $x_f^r$  using  $a_{MAP}$  by computing the posterior probability of  $X_f$  as shown in Fig 3f. If this probability of reaching the desired state is above a prespecified threshold,  $\tau$ , the robot executes  $a_{MAP}$ , otherwise it executes the "Ask human" action to request human assistance.

## Robot

We use the Gambit robot arm-and-gripper (Fig 2a) designed at Intel Labs Seattle. Gambit is well-suited to tabletop manipulation tasks with small objects and has previously been shown to perform well in tasks with humans in the loop [18].

**Sensing.** For sensing the current state of objects on the table, for example, during human demonstrations, we use the Microsoft Kinect RGBD camera. The Kinect is mounted on the base frame of Gambit and looks down on the table surface. The robot takes as input the stream of RGB and depth images from the Kinect and first segments out the background and the hand of the human holding the small objects. The remaining pixels are then used to determine the state of objects on the table using a simple heuristic based on centroids. We define three discrete areas that are used for defining the state of the objects as shown in Fig 2a—(1) "LEFT" signifying that the object is on the left side of the blueline on the table; (2) "RIGHT" denoting that the object is on the right side of the blueline; and (3) "OFFTABLE" signifying that the object is not in the tabletop work area.

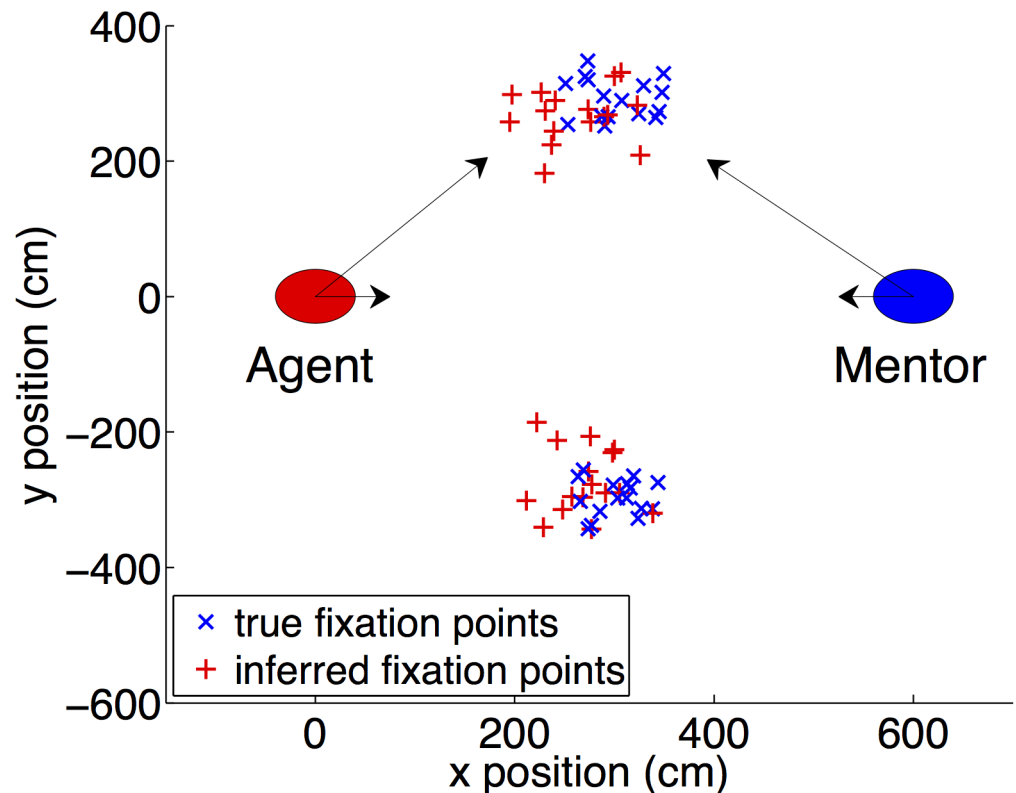
**Robot Action.** We assume that the robot possesses a fixed set of six high-level actions for manipulating objects: place LEFT (PlaceL), place RIGHT (PlaceR), place OFFTABLE (PlaceOt), push to LEFT (pUshL), push to RIGHT (pUshR), and push OFFTABLE (pUshOt). For the "place" actions, the robot first attempts to pick up the object by moving its end effector above the centroid of the object and rotating the gripper to align itself perpendicular to the major axis of the object. If the robot successfully picked up the object, it places the object down at the location (LEFT, RIGHT, or OFFTABLE) indicated by the place command. For the "push" actions, the robot first positions its end effector behind the object based on its centroid and direction of the push. For pUshL and pUshR, the gripper yaw angle is rotated perpendicular to the major axis of the table, while for pUshOt, it is rotated parallel to the major axis of the table. This ensures that object contact area is maximized to reduce the chance of the object slipping while pushing. The robot pushes the object until it changes state (or the inverse kinematics solver fails to find a possible solution).

## Results

### Gaze Following: Model Simulations

To test our model, we randomly sample goal positions and then compute the action required to fixate on this goal. We add Gaussian noise to this action, compute the resulting gaze vector if



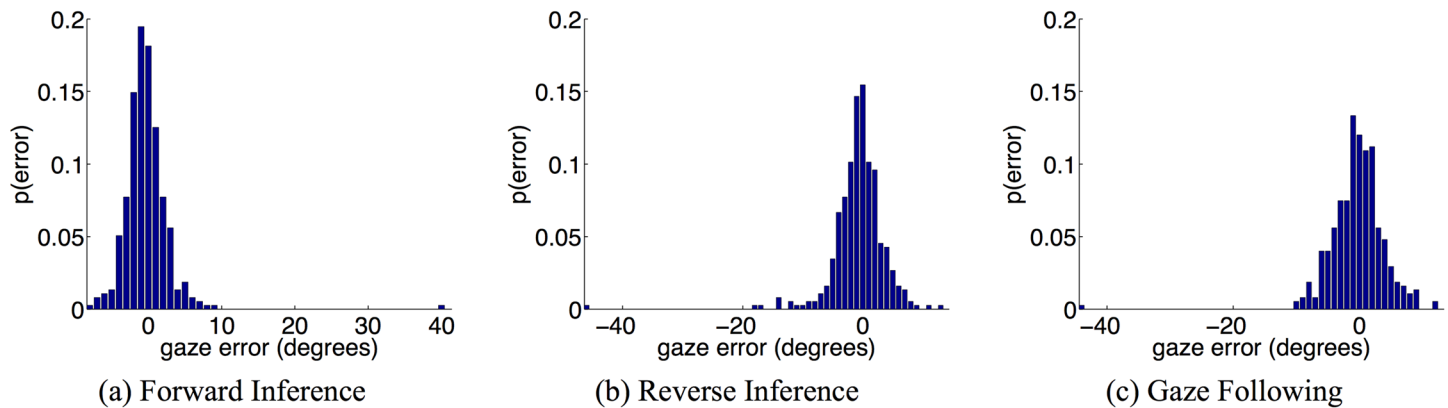


**Fig 4. Gaze following results.** The agent and mentor face each other in a 2D simulated environment. Goal positions inferred by the model from mentor observations are shown in red next to true goal positions (blue) for sampled goal positions both to the left and to the right of the agent. Black arrows represent the initial and final gaze vectors of the agent and mentor for one of these test data points. In this simple example, the goal locations were equidistant from the agent and mentor but the model readily generalizes to other cases as well.

doi:10.1371/journal.pone.0141965.g004

this action were taken, and add Gaussian noise to this gaze vector. This method is equivalent to training the model with rejection sampling wherein the agent rejects all samples that do not result in successful fixation on the goal position. The Gaussian processes are trained on this randomly generated data and then tested on separate test data. The default reference frame for both agent and mentor is at the origin gazing along the x-axis. Each agent has their own reference frame and we assume that we know the transformation from the mentor's reference frame to the agent's. This transformation is not learned by our model but we believe that this is a minor assumption, especially since we already assume the agent can observe the mentor's position and head pose.

The mentor and agent are positioned as shown in Fig 4. Goal locations for the training data were generated uniformly at random from the area between the agent and the mentor (within the rectangle formed by  $x$  in  $[100, 500]$  and  $y$  in  $[-500, 500]$ , where the agent is at  $(0,0)$  and the mentor is at  $(600,0)$ ). We used Gaussian noise with standard deviation of 3 degrees for angles and a standard deviation of 10 cm for locations and distances. For reverse inference, the prior goal state  $P(G)$  is a Gaussian centered halfway between the two agents along the x-axis. While this prior is quite weak, a single observation of  $(x_i, x_f)$  is insufficient to overcome the prior in reverse inference. Instead, we use a single observation of  $x_i$  and five observations of  $x_f$  to get an



**Fig 5. Accuracy of goal inference and gaze following.** Histogram plots showing the probability of an error (in degrees) between the inferred and the true gaze vector—the gaze vectors in the final state  $X_f$  were used for (a) and (c), and the gaze vectors in the the goal state  $G$  were used for (b). These probabilities were estimated from 375 test points spread uniformly over the test region. Note how the accuracy gracefully decays as more complicated inference is performed ((a) is the simplest, while (c) is the most complex).

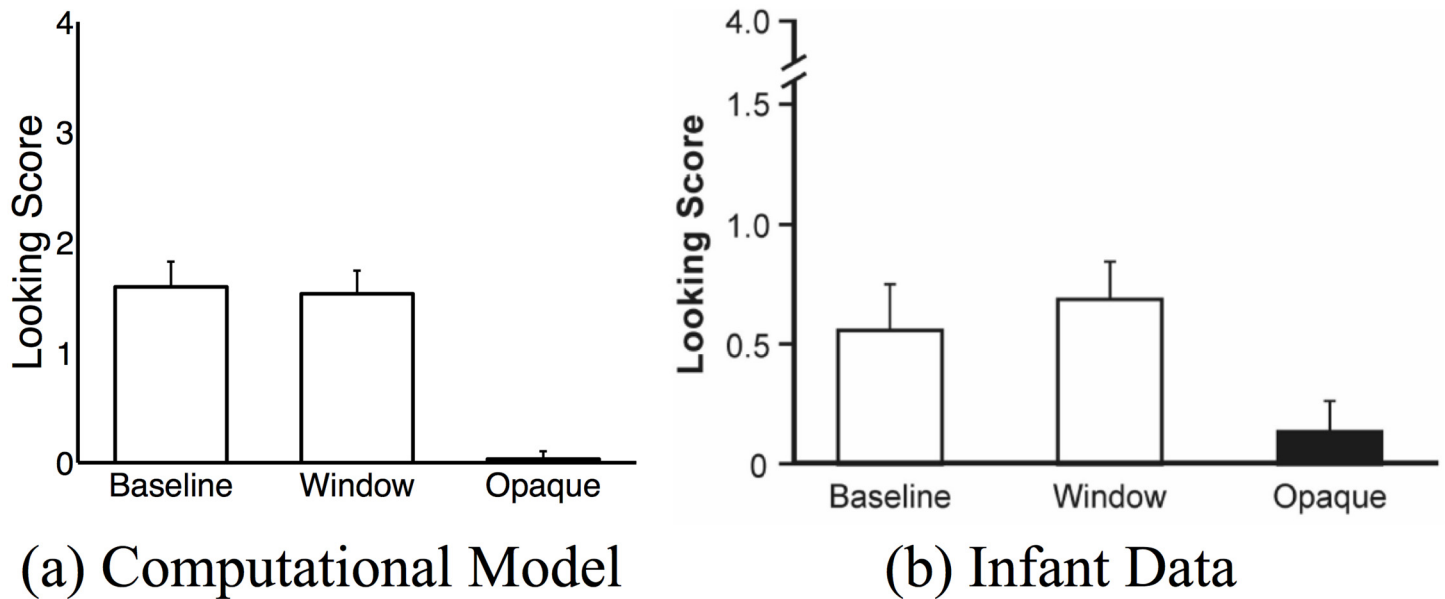
doi:10.1371/journal.pone.0141965.g005

accurate estimate of the goal distribution  $P(G|x_i, x_f)$ . More precisely, we run reverse inference with the observed values  $(x_i, x_f^{(1)})$  to compute  $P(G | x_i, x_f^{(1)})$ , and then use this as the prior for a second run of reverse inference to compute  $P(G | x_i, x_f^{(1:2)})$ . We repeat this five times to compute  $P(G | x_i, x_f^{(1:5)})$ . We believe such an inference process could be executed within the short amount of time taken for gaze following.

**Gaze Following: Model Performance.** We found that the model learns accurate transition and policy functions from small amounts of noisy training data ( $n = 200$  data points in our accuracy tests). The nonparametric nature of Gaussian processes ensured very little customization was required. The model runs in sub-minute times for the dimensionality we are using, though additional approximations may be necessary to scale well to high dimensions. We also found that the Gaussian approximations we are making have little effect because the data is generally unimodal and close to symmetric.

Figs 4 and 5 show performance results for the model as it performs forward inference, reverse inference, and gaze following (combined reverse and forward inference). The model is robust to noise and is able to provide accurate gaze following results even though additional levels of uncertainty are introduced by the second level of inference.

**Blindfold Self-Experience Task.** In order to test the cognitive plausibility of our model, we recreated the experiments from [6], where infants' self-experience with a blindfold affects whether or not they follow the gaze of blindfolded adults (see Fig 1c). We trained 60 separate agents with our model on randomly generated training data. Similar to the infant experiments, one third of these agents were given additional experience with a blindfold wherein they train an additional  $GP_\pi$  for their model so that they now have  $GP_\pi$  and  $GP_\pi^B$ , where  $GP_\pi$  is the original GP of the model and  $GP_\pi^B$  is the GP with blindfold experience. The agents with the blindfold experience learn the consequences of wearing an opaque blindfold in  $GP_\pi^B$ . The other 40 agents were trained normally although one group is the baseline group and the other is the windowed blindfold group. In our simulations, the windowed blindfold group experiences the same training data as the no-blindfold case since, in both cases, the overall result is that the target is visible to the agent. So these two groups will be identical except for noise.



**Fig 6. Comparison of the model to infant data from [6].** Error bars represent standard error of means.

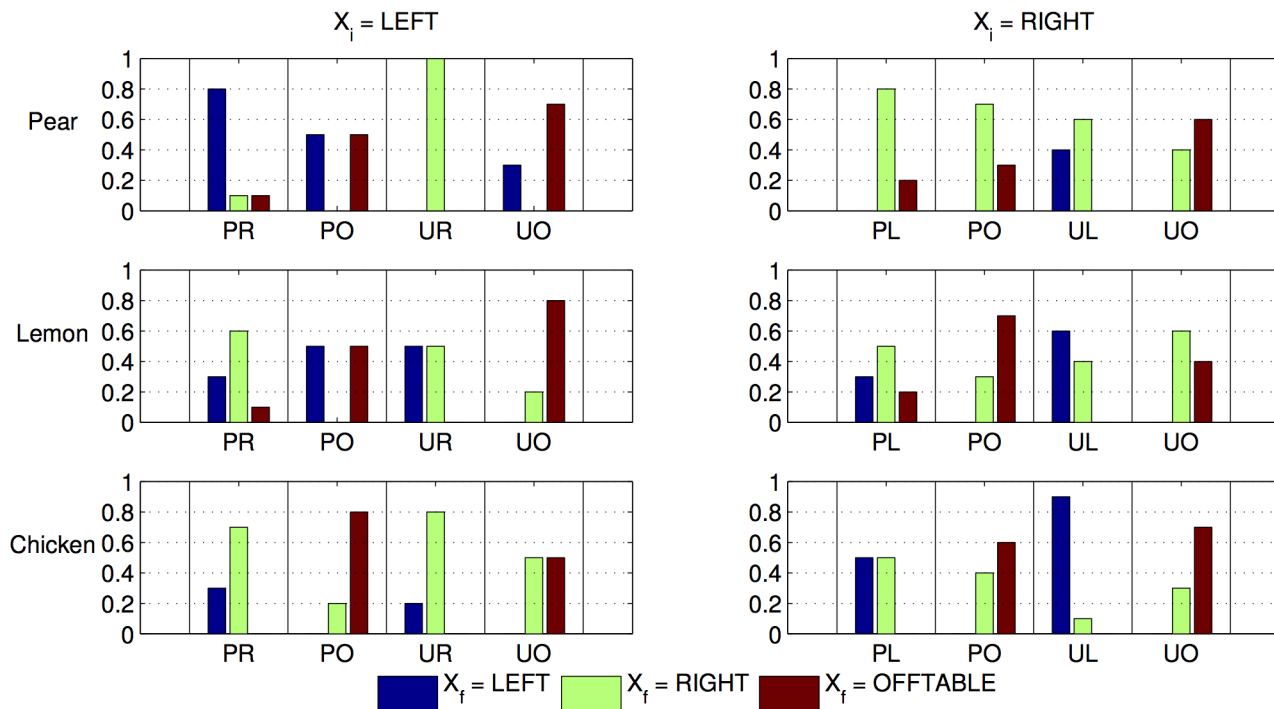
doi:10.1371/journal.pone.0141965.g006

Each agent is presented with 4 trials where it observes a mentor make a head turn to face either 45 degrees to the left or 45 degrees to the right (plus noise). Trials are scored as +1 if the agent turns its head at least 30 degrees in the direction of the correct target and -1 if it turns its head at least 30 degrees in the direction of the wrong target. The agents with no blindfold experience use the basic model (which contains no blindfold knowledge) and thus assume that the mentor is fixating on an object to the left or to the right. Those with blindfold experience observe that the mentor is wearing a blindfold and use their learned  $GP_{\pi}^B$  for the reverse inference (applying their model to the mentor). For this group, we expect the agent to have learned through self-experience that when blindfolded, there is no correlation between a head movement and any particular goal position. The blindfold-experienced agent then uses  $GP_{\pi}$  for the forward inference (because the agent is not wearing a blind-fold). In order to simulate infants with little experience with blindfolds, we provided the agents in this experiment with very little training data ( $n = 15$ ). If we use more training data, the agents perform almost perfectly in this task.

Results are shown in Fig 6 and match the pattern of results in Meltzoff and Brooks (reproduced in Fig 6a for comparison). This suggests that gaze following involves understanding the underlying intention or goal of the mentor, and that self-experience plays a major role in learning the consequences of intentions and related actions.

### Robotic Tabletop Manipulation Task

To illustrate the discrete-valued version of the model, we used a tabletop manipulation with three toy objects of different shapes: a “pear,” a “lemon,” and a miniature “broiled chicken.” These objects could be in one of three different states: LEFT, RIGHT, and OFFTABLE, as defined in the previous section. The robot’s aim is to learn the consequences of actions on these objects, infer the goals of human actions, and imitate the actions of humans manipulating these objects on the table.



**Fig 7. Robotic task: Learned transition model.** Each row represents different types of objects, and each column represents the different initial states  $X_i$ . The colors of bars represent different final states  $X_f$ . The y-axis represents the range of probabilities and the x-axis represents the six different manipulation actions available to the robot (PL = place LEFT, PR = place RIGHT, PO = place OFFTABLE, UL = pUsh LEFT, UR = pUsh RIGHT, UO = pUsh OFFTABLE). We do not show actions that cause self-transitions given an initial state.

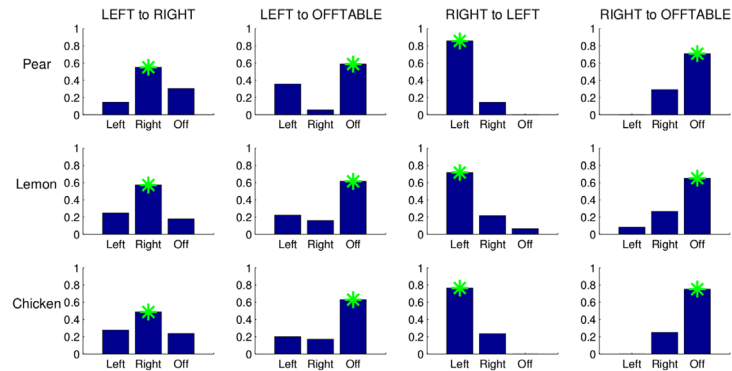
doi:10.1371/journal.pone.0141965.g007

Fig 7 shows the learned transition probabilities for manipulating these objects based on the robot’s self-discovery phase. The transition models are learned by performing 10 trials for each initial state and action pair  $(x_i, a)$  for each object type. We deliberately used a small number of trials to test whether the method could cope with less training data and more uncertainty in the transition model.

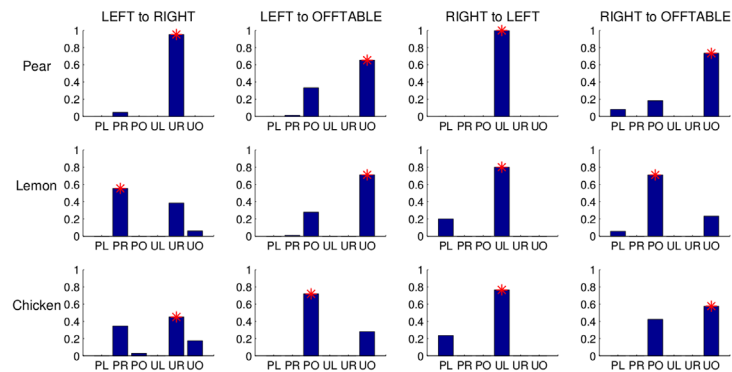
Since the state space is small, we are able to enumerate and test all of the interesting and possible human demonstrations. By interesting, we mean that the state changes after action execution (for example, from RIGHT to OFFTABLE) and the initial state is not OFFTABLE. There are in total four interesting state changes for each object that can be demonstrated by a human: LEFT to RIGHT, LEFT to OFFTABLE, RIGHT to LEFT, and RIGHT to OFFTABLE. Note that our current implementation does not allow the robot to pick up objects that are located OFFTABLE

Fig 8a shows the inferred goals given all possible interesting state changes, using the graphical model in Fig 3d. For all cases, our model correctly infers the intended goal state of the human. Fig 8b shows the maximum a posteriori probability (MAP) action for a given initial state and goal. Our model correctly identifies whether a “place” or a “push” action is better, given the dynamics of the object as encoded in the learned probabilistic model for the object. Note that the two most preferred actions are always push or place actions in the correct direction.

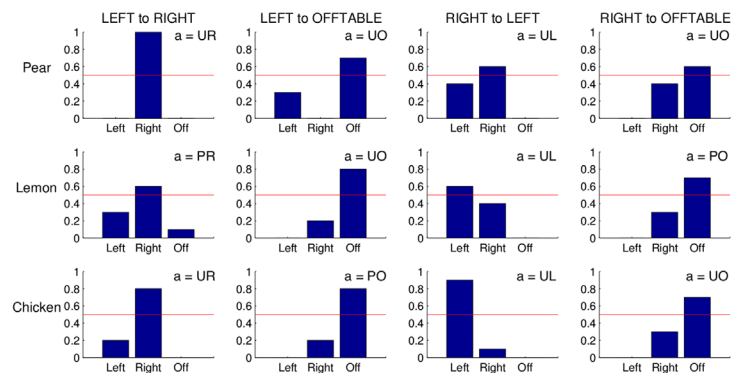
Finally, Fig 8c shows the predicted state distribution given an action and an initial state. The robot calculates the posterior probability of getting to the desired state, and executes the action



(a) Forward Inference



(b) Action Inference



(c) Final State Prediction

**Fig 8. Robotic task: Results.** (a) **Most likely goals:** Initial and final states are at the top of each column. The height of the bar represents the posterior probability of each goal state, with the true goal state marked by an asterisk. (b) **Inferring actions:** For each initial and desired final state, the plots show the posterior probability of each of the six actions, with the MAP action indicated by an asterisk. (c) **Predicting final state:** The plots show the posterior probability of reaching the desired final state, given the initial state and the corresponding MAP action shown in (b). The red bar marks 0.5, the threshold below which the robot asks for human help in the Interactive Goal-Based mode.

doi:10.1371/journal.pone.0141965.g008

**Table 1. Comparison of approaches.**

	Trajectory	Goal-Based	Interactive Goal-Based
Pick & Place Demonstration			
Pear	0.2250	0.6750	<i><b>0.8250</b></i>
Lemon	0.5250	<b>0.6750</b>	<b>0.6750</b>
Chicken	0.6500	<b>0.8000</b>	<b>0.8000</b>
Push Demonstration			
Pear	0.6750	0.6750	<i><b>0.7250</b></i>
Lemon	0.5750	<b>0.6750</b>	<b>0.6750</b>
Chicken	0.7250	<b>0.8000</b>	<b>0.8000</b>

The hypothetical success rates of three different approaches to imitation are shown. The success rate shown in each cell is averaged over four interesting state changes mentioned in the main text. The boldfaced results indicate the row-wise highest success rates, and the cases where the robot’s request for human help improved success the rates are shown in italic.

doi:10.1371/journal.pone.0141965.t001

if this probability is above a predetermined threshold (%50). Otherwise, it asks the human collaborator. For example, the predicted output of moving a pear from RIGHT to LEFT (1st row of 3rd column in Fig 8c) using the pushing left action is lower than %50, and therefore the robot will ask request the human help.

Table 1 compares “trajectory-based” imitation of the human demonstration with our proposed goal-based approach. The trajectory-based approach simply mimics the human action without considering the goal or uncertainty, i.e., it executes a place action if the human executes a place, and a push action if the human executes a push. The goal-based approach on the other hand recognizes the goal and uses the best action it has available to achieve the goal.

Using our computed transition probabilities, we can calculate the hypothetical success rate of a purely trajectory-based approach. For our goal-based approach, we use the posterior distribution shown in Fig 8b. Finally, the “Interactive Goal-Based” mode assumes that the robot may ask a human for help, with a 100% success rate when the human executes the requested action. The third column in Table 1 shows what the performance would be if we require the robot to be 50% sure of reaching a desired state. We do not see perfect imitation results on the third column because the robot does not ask the human for help in every case. In some cases, the probability of success will surpass the confidence threshold, but the goal state may not be reached after the action is executed.

The results demonstrate the expected behavior of the goal-based method. The success rates in goal-based method (Table 1, second column) are identical across the different actions demonstrated (e.g. pick & place row vs. push row). In addition, the results demonstrate the advantage of a goal-based approach over purely trajectory-based imitation, and its potential as a human-robot collaboration mechanism.

### Contributions and Comparisons with Related Work

Our work makes contributions both to the fields of robotics and cognitive modeling. In robotics, our approach contributes to the growing number of efforts leveraging biological paradigms and developmental science to design new methods for social robotics and human-robot interaction [17, 19–30]. Specifically, our proposed framework for goal inference and imitation lays



the foundation for a new approach to designing social robots as robots that learn internal models of their environment through self-experience and utilize these models for human intent recognition, skill acquisition from human observation, and human-robot collaboration. Although our results are based on proof-of-concept systems, the underlying Bayesian framework is general. We anticipate being able to progressively increase the sophistication of our models by leveraging the rapid advances being made in probabilistic reasoning, Bayesian modeling, and learning. As an illustrative example, the success of simultaneous localization and mapping (SLAM) algorithms in robotics [31] in the past decade can be largely attributed to the availability of efficient algorithms and computing platforms for probabilistic reasoning, within the broader context of probabilistic robotics [32].

Our approach to robotic imitation emphasizes the importance of goal inference in imitation learning, compared to traditional methods for programming-by-demonstration that have relied on following action trajectories. Previous approaches to robotic imitation that have also relied on high-level goals (e.g., [33, 34]) have not emphasized self-discovery of probabilistic models, a central tenet of the developmental approach proposed here for bootstrapping goal-based imitation. Other robotics work has focused on attempting to model continuous low-level goals [35]. Neither of these approaches adopt probabilistic models for human-robot interaction tasks. Approaches that do utilize Bayesian methods for goal inference [36, 37] and robotic imitation [38] have done so without the developmental science perspective that we bring to the problem.

In the context of cognitive modeling, our model can be regarded as a Bayesian instantiation of the “Like-Me” developmental hypothesis [7, 39]. It acknowledges a role for a certain type of mental simulation [8] as well as a learned “theory” [9, 10] of actions and their consequences using a probabilistic model learned from experience (see also [40]). The model is closely related to the goal-based imitation model of Verma and Rao [3, 17, 41] and the inverse planning model of Baker et al. [42]. Also related to our approach are models for planning based on probabilistic inference [3, 43–46].

Meltzoff’s “Like-Me” hypothesis has previously been applied in robotics to tackle the important problem of “whom to imitate” [47]: In this case, a robot first builds a self model, and then uses this self-model to discover self-other equivalences with other robots so as to distinguish an appropriate from an inappropriate “mentor” robot. Note that in this case as well as in our own work, the “mentor” agent is not necessarily an active teacher but is only being observed (cf. [48]). A recent developmental approach to robotic imitation [49] also focuses on tabletop manipulation tasks, with greater complexity than our experiments, but without the benefits of bootstrapping derived from the “Like-Me” hypothesis inherent in our framework.

## Summary and Conclusion

Our results suggest that the process of imitation-based learning can be bootstrapped by (i) learning a probabilistic model of the perceptual consequences of one’s own actions through self-experience, (ii) using this learned model to infer the goals of actions of others, and (iii) using the inferred goals to perform goal-based imitation. Such a strategy works even when the imitator’s actuators are different from the demonstrator’s. We first showed how such a model can emulate infant development of gaze following by implementing gaze following via goal inference in learned graphical models based on Gaussian processes. Using a table-top robotic system, we demonstrated that a robot can learn probabilistic models of actions on objects and utilize these models for inferring the intent of human actions on objects. Additionally, we showed that the approach facilitates human-robot collaboration by allowing the robot to predict the success probability of achieving a goal and to seek human help when success probability is low.

Our results point to a number of interesting open issues. For the studies in this paper, we used a simple exhaustive exploration strategy, but for larger state spaces, a more sophisticated approach based on reward functions (e.g., [50]) could be employed. Additionally, following the example of human infants, some form of directed self-exploration based on observing human mentors (e.g., [51]) may be desirable. The model assumes that states in the environment are known (corresponding to the case of MDPs or Markov decision processes)—an interesting direction for future work is extending the model to the more realistic case where only observations of states are available (partially observable MDPs or POMDPs) and where learning involves optimizing rewards. Finally, our current model involves inference over a single time-step, which simplifies the inference problem. We hope to explore the applicability of our model for multi-step inference problems in future work.

The approach we have presented lends itself naturally to generalization based on relational probabilistic models [52, 53] and hierarchical Bayesian representations [54]. Such models have the potential to significantly increase the scalability and applicability of our suggested approach to large-scale scenarios, besides facilitating transfer of learned skills across tasks and domains. We intend to investigate such relational models in future work.

## Acknowledgments

This research was supported by ONR Science of Autonomy grant no. N000141310817, ONR Cognitive Science program grant no. N000140910097, NSF grant no. SMA-1540619 and NSF grant no. IIS-1318733, by the Intel Science and Technology Center (ISTC), Seattle, and by an NSERC fellowship to A. Friesen. The authors would like to thank Marc Deisenroth, Cynthia Matuszek and Kendall Lowrey for their advice, sharing code, and for help during the initial stages of this project.

## Author Contributions

Conceived and designed the experiments: RR MC AF AM DF. Performed the experiments: MC AF. Analyzed the data: MC AF. Contributed reagents/materials/analysis tools: MC AF. Wrote the paper: RR MC AF AM DF.

## References

1. Billard A, Calinon S, Dillmann R, Schaal S. Robot programming by demonstration. In: Springer handbook of robotics. Springer; 2008. p. 1371–1394.
2. Grimes D, Chalodhorn R, Rao R. Dynamic imitation in a humanoid robot through nonparametric probabilistic inference. In: Proceedings of Robotics: Science and Systems; 2006.
3. Verma D, Rao R. Goal-based imitation as probabilistic inference over graphical models. In Advances in neural information processing systems. 2006; 18:1393–1400.
4. Meltzoff AN, Kuhl PK, Movellan J, Sejnowski TJ. Foundations for a new science of learning. Science. 2009; 325:284–288. doi: [10.1126/science.1175626](https://doi.org/10.1126/science.1175626) PMID: [19608908](https://pubmed.ncbi.nlm.nih.gov/19608908/)
5. Meltzoff AN. Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. Developmental Psychology. 1995; 31(5):838–850. doi: [10.1037/0012-1649.31.5.838](https://doi.org/10.1037/0012-1649.31.5.838) PMID: [25147406](https://pubmed.ncbi.nlm.nih.gov/25147406/)
6. Meltzoff AN, Brooks R. Self-experience as a mechanism for learning about others: A training study in social cognition. Developmental Psychology. 2008; 44(5):1257–1265. doi: [10.1037/a0012888](https://doi.org/10.1037/a0012888) PMID: [18793060](https://pubmed.ncbi.nlm.nih.gov/18793060/)
7. Meltzoff AN. 'Like me': A foundation for social cognition. Developmental Science. 2007; 10(1):126–134. doi: [10.1111/j.1467-7687.2007.00574.x](https://doi.org/10.1111/j.1467-7687.2007.00574.x) PMID: [17181710](https://pubmed.ncbi.nlm.nih.gov/17181710/)
8. Davies M, Stone T. Mental simulation: Evaluations and applications. Blackwell Oxford; 1995.
9. Churchland PM. Folk psychology and the explanation of human behavior. Philosophical Perspectives. 1989; 3:225–241. doi: [10.2307/2214269](https://doi.org/10.2307/2214269)
10. Gopnik A, Meltzoff AN. Words, thoughts, and theories. MIT Press; 1997.

11. Meltzoff AN, Moore MK. Explaining facial imitation: A theoretical model. *Early Development and Parenting*. 1997; 6(34):179–192.
12. Rasmussen CE. Gaussian processes for machine learning. 2006;.
13. Rasmussen CE, Ghahramani Z. Bayesian monte carlo. In: *Advances in neural information processing systems*; 2003. p. 489–496.
14. Scherberger H, Andersen RA. Sensorimotor transformation in the posterior parietal cortex. *The Visual Neurosciences*. 2003;p. 1324–1336.
15. Sauser EL, Billard AG. Three-dimensional frames of references transformations using recurrent populations of neurons. *Neurocomputing*. 2005; 64:5–24. doi: [10.1016/j.neucom.2004.11.019](https://doi.org/10.1016/j.neucom.2004.11.019)
16. Sauser EL, Billard AG. Interferences in the transformation of reference frames during a posture imitation task. In: *Artificial Neural Networks–ICANN 2007*. Springer; 2007. p. 768–778.
17. Rao RPN, Shon AP, Meltzoff AN. A Bayesian model of imitation in infants and robots. In: *Imitation and social learning in robots, humans, and animals*. 2007;p. 217–247. doi: [10.1017/CBO9780511489808.016](https://doi.org/10.1017/CBO9780511489808.016)
18. Matuszek C, Mayton B, Aimi R, Deisenroth MP, Bo L, Chu R, et al. Gambit: An autonomous chess-playing robotic system. In: *2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE; 2011. p. 4291–4297.
19. Mataric MJ. Behavior-based robotics as a tool for synthesis of artificial behavior and analysis of natural behavior. *Trends in cognitive sciences*. 1998; 2(3):82–86. doi: [10.1016/S1364-6613\(98\)01141-3](https://doi.org/10.1016/S1364-6613(98)01141-3) PMID: [21227083](https://pubmed.ncbi.nlm.nih.gov/21227083/)
20. Metta G, Sandini G, Konczak J. A developmental approach to visually-guided reaching in artificial systems. *Neural networks*. 1999; 12(10):1413–1427. doi: [10.1016/S0893-6080\(99\)00070-2](https://doi.org/10.1016/S0893-6080(99)00070-2) PMID: [12662625](https://pubmed.ncbi.nlm.nih.gov/12662625/)
21. Lungarella M, Metta G, Pfeifer R, Sandini G. Developmental robotics: a survey. *Connection Science*. 2003; 15(4):151–190. doi: [10.1080/09540090310001655110](https://doi.org/10.1080/09540090310001655110)
22. Asada M, MacDorman KF, Ishiguro H, Kuniyoshi Y. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems*. 2001; 37(2):185–193. doi: [10.1016/S0921-8890\(01\)00157-9](https://doi.org/10.1016/S0921-8890(01)00157-9)
23. Breazeal C, Scassellati B. Challenges in building robots that imitate people. In: *Imitation in Animals and Artifacts*. 2002;.
24. Breazeal C, Buchsbaum D, Gray J, Gatenby D, Blumberg B. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life*. 2005; 11(1–2):31–62. doi: [10.1162/1064546053278955](https://doi.org/10.1162/1064546053278955) PMID: [15811219](https://pubmed.ncbi.nlm.nih.gov/15811219/)
25. Scassellati B. Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot. In: *Computation for Metaphors, Analogy, and Agents*. Springer; 1999. p. 176–195.
26. Billard A, Mataric MJ. A biologically inspired robotic model for learning by imitation. In: *Proceedings of the fourth international conference on Autonomous agents*. ACM; 2000. p. 373–380.
27. Schaal S. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*. 1999; 3(6):233–242. doi: [10.1016/S1364-6613\(99\)01327-3](https://doi.org/10.1016/S1364-6613(99)01327-3) PMID: [10354577](https://pubmed.ncbi.nlm.nih.gov/10354577/)
28. Meltzoff AN, Brooks R, Shon AP, Rao RP. “Social” robots are psychological agents for infants: A test of gaze following. *Neural Networks*. 2010; 23:966–972. doi: [10.1016/j.neunet.2010.09.005](https://doi.org/10.1016/j.neunet.2010.09.005) PMID: [20951333](https://pubmed.ncbi.nlm.nih.gov/20951333/)
29. Copete JL, Nagai Y, Asada M. Development of goal-directed gaze shift based on predictive learning. In: *Development and Learning and Epigenetic Robotics (ICDL-Epirob)*, 2014 Joint IEEE International Conferences. IEEE; 2014. p. 334–339.
30. Cangelosi A, Schlesinger M. *Developmental robotics: From babies to robots*. MIT Press; 2015.
31. Montemerlo M, Thrun S, Koller D, Wegbreit B. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In: *AAAI/IAAI*; 2002. p. 593–598.
32. Thrun S, Burgard W, Fox D. *Probabilistic robotics*. MIT press; 2005.
33. Song D, Huebner K, Kyrki V, Kragic D. Learning task constraints for robot grasping using graphical models. In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE; 2010. p. 1579–1585.
34. Ek CH, Song D, Huebner K, Kragic D. Task modeling in imitation learning using latent variable models. In: *10th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2010. IEEE; 2010. p. 548–553.

35. Wang Z, Mülling K, Deisenroth MP, Amor HB, Vogt D, Schölkopf B, et al. Probabilistic movement modeling for intention inference in human–robot interaction. *The International Journal of Robotics Research*. 2013; 32(7):841–858. doi: [10.1177/0278364913478447](https://doi.org/10.1177/0278364913478447)
36. Patel M, Ek CH, Kyriazis N, Argyros A, Miro JV, Kragic D. Language for learning complex human-object interactions. In: 2013 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2013. p. 4997–5002.
37. Kwon WY, Suh IH. A temporal bayesian network with application to design of a proactive robotic assistant. In: 2012 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2012. p. 3685–3690.
38. Kroemer O, Daniel C, Neumann G, van Hoof H, Peters J. Towards Learning Hierarchical Skills for Multi-Phase Manipulation Tasks. In: 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2015.
39. Meltzoff AN. The 'like me' framework for recognizing and becoming an intentional agent. *Acta Psychologica*. 2007; 124(1):26–43. doi: [10.1016/j.actpsy.2006.09.005](https://doi.org/10.1016/j.actpsy.2006.09.005) PMID: [17081488](https://pubmed.ncbi.nlm.nih.gov/17081488/)
40. Waismeyer A, Meltzoff AN, Gopnik A. Causal learning from probabilistic events in 24-month-olds: An action measure. *Developmental Science*. 2015; 18(1):175–182. doi: [10.1111/desc.12208](https://doi.org/10.1111/desc.12208) PMID: [25041264](https://pubmed.ncbi.nlm.nih.gov/25041264/)
41. Shon AP, Storz JJ, Meltzoff AN, Rao RP. A cognitive model of imitative development in humans and machines. *International Journal of Humanoid Robotics*. 2007; 4(02):387–406. doi: [10.1142/S0219843607001059](https://doi.org/10.1142/S0219843607001059)
42. Baker CL, Saxe R, Tenenbaum JB. Action understanding as inverse planning. *Cognition*. 2009; 113(3):329–349. doi: [10.1016/j.cognition.2009.07.005](https://doi.org/10.1016/j.cognition.2009.07.005) PMID: [19729154](https://pubmed.ncbi.nlm.nih.gov/19729154/)
43. Verma D, Rao RP. Planning and acting in uncertain environments using probabilistic inference. In: *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference*. IEEE; 2006. p. 2382–2387.
44. Toussaint M, Storkey A. Probabilistic inference for solving discrete and continuous state Markov Decision Processes. In: *Proceedings of the 23rd international conference on Machine learning*. ACM; 2006. p. 945–952.
45. Verma D, Rao RP. Imitation learning using graphical models. In: *Machine Learning: ECML 2007*. Springer; 2007. p. 757–764.
46. Botvinick M, An J. Goal-directed decision making in prefrontal cortex: A computational framework. In: *Advances in Neural Information Processing Systems*; 2009. p. 169–176. PMID: [25258502](https://pubmed.ncbi.nlm.nih.gov/25258502/)
47. Kaipa KN, Bongard JC, Meltzoff AN. Self discovery enables robot social cognition: Are you my teacher? *Neural Networks*. 2010; 23(8):1113–1124. doi: [10.1016/j.neunet.2010.07.009](https://doi.org/10.1016/j.neunet.2010.07.009) PMID: [20732790](https://pubmed.ncbi.nlm.nih.gov/20732790/)
48. Hoppitt WJ, Brown GR, Kendal R, Rendell L, Thornton A, Webster MM, et al. Lessons from animal teaching. *Trends in Ecology & Evolution*. 2008; 23(9):486–493. doi: [10.1016/j.tree.2008.05.008](https://doi.org/10.1016/j.tree.2008.05.008)
49. Ugur E, Nagai Y, Sahin E, Oztop E. Staged Development of Robot Skills: Behavior Formation, Affordance Learning and Imitation with Motionese. *IEEE Transactions on Autonomous Mental Development*. 2015; 7:119–139. doi: [10.1109/TAMD.2015.2426192](https://doi.org/10.1109/TAMD.2015.2426192)
50. Deisenroth MP, Rasmussen CE. PILCO: A model-based and data-efficient approach to policy search. In: *Proceedings of the 28th International Conference on Machine Learning*; 2011.
51. Grimes DB, Rao RP. Learning actions through imitation and exploration: Towards humanoid robots that learn from humans. In: *Creating Brain-Like Intelligence*. Springer; 2009. p. 103–138.
52. Koller D, Friedman N. *Probabilistic graphical models: principles and techniques*. MIT Press; 2009.
53. Richardson M, Domingos P. Markov logic networks. *Machine Learning*. 2006; 62(1):107–136. doi: [10.1007/s10994-006-5833-1](https://doi.org/10.1007/s10994-006-5833-1)
54. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. How to grow a mind: Statistics, structure, and abstraction. *Science*. 2011; 331(6022):1279–1285. doi: [10.1126/science.1192788](https://doi.org/10.1126/science.1192788) PMID: [21393536](https://pubmed.ncbi.nlm.nih.gov/21393536/)